■論　文■

Costello: Ongoing Experiments in an Internet Virtual Reality

Adrian Cohen

In an earlier paper, (Cohen, 2000) I introduced the concept of a text-based internet virtual reality (commonly called a MUD), as a pedagogical tool for use in an EFL setting. It was suggested that the interactive nature of such environments could provide motivation and opportunities for students to communicate in English, while the text basis of the virtual world would provide for comprehensible input according to the Krashen and Terrell (1983) model. It was also hypothesised that the use of a corrective language parser to process student input into the computer could help promote the acquisition of correct language forms as suggested by Ellis (1994). The virtual reality in question ('Costello') being a game, it was argued that the motivational benefits of the system would outweigh the technological barriers that students would have to overcome, allowing them to achieve similar or better results in a more enjoyable way than could be provided with traditional homework tasks. The current paper has two purposes: to report on the current developmental status of the Costello virtual reality, and to detail the results of an initial experimental study comparing a class using the Costello RPG with a class using a conventional extensive reading program.

At the time of the previous paper, the Costello MUD game was undergoing a fundamental overhaul aimed at dramatically increasing the functionality and pedagogical effectiveness of the game. These changes involved rewriting almost all aspects of the software and documentation associated with the virtual reality, a database of files and programs that surpasses some 20,000 in number. The main aims of this overhaul were to provide a language parser that could successfully understand almost any language that users would reasonably be expected to use in the game, the development of a large world and stimulating tasks (called 'quests') that would provide a rich linguistic experience for students, and the implementation of software that would allow students to undertake creative writing by adding their own areas to the virtual world. In the first part of this paper I will give a brief summary of the changes that have been completed and their effectiveness (or otherwise) in providing a stimulating environment for the students.

Part One: Costello Updated

The parser

The language parser, and its ability to understand natural English, are essential to the pedagogical goals of the project. It is this that allows Costello to escape the criticisms referred to in Cohen (2000) that such games might promote the acquisition of 'telegraphese' by students (Higgins and Johns, 1984). While the original version of Costello could understand sentences using the articles 'a' and 'the' (itself a surprisingly rare quality in such games), it was unable to go much beyond the understanding of simple sentences such as 'Take the bag.' The current version of the parser goes far beyond such limitations and can process a lot of natural language that fits the register of the virtual reality world. Examples of more complicated sentences that the game can now understand include:

Drop everything except the bag, helmet and lantern.
Pick up the blue umbrella and go south.
Take the key and put the key in the bag

Take the key from the bag and unlock the door with the key
Put the red ball and the blue ball in the bag and give the bag to John

The parser can also now resolve simple ambiguities by asking a confirmation question:

Put the ball in the bag.
There is more than one ball here. Which ball?
The blue one.

However, it was found that this degree of sophistication was too high for the level of students currently using the software (first year university students taking compulsory English classes). Very few students were able to realise that they had entered an ambiguous sentence, and still fewer were able to resolve the ambiguity by answering the computer's question. As a result, this feature has now been disabled and the computer simply picks the first item in a list of items that the student is currently carrying in the virtual world.

There are still a number of limitations that need to be addressed in the parser before it can be considered completely satisfactory. For example, it cannot handle referential pronouns in sentences such as 'Take the key from the bag and unlock the door with it.' Nevertheless, the new parser is a great advance on the previous system and includes both grammar correction and a scoring system which gives students more points for entering longer commands to the game.

The Virtual World and Quests

The Costello virtual world has been expanded enormously during the past year and represents a far richer textual environment than before. The world is now so large that there are over 30,000 locations that players can visit, although it should be pointed out that many of these are repetitive (when you are lost in the forest, one part of the forest looks much like another). This represents a lot of reading for the students as they move around the world, and they are now rewarded for exploring new areas that are not repetitive in nature by gaining game points as they move around. This reflects the fact that in order to find their way around the world, they must read the descriptions of the places in which they travel. However, a criticism that can be made here is that it is possible for lazy students to make a map of the Costello world without reading any of the English descriptions at all; all they need do is note the directions in which they move. While this is true, there are a number of ways in which this temptation is reduced. Firstly, and perhaps most importantly, simply moving around the world, without taking any time to read the descriptions, is a very boring way to pass one's time. Secondly, students are required each week to write a report of their travels in the virtual world (using their own words) and of their adventures in it, and finally it is impossible to complete any of the quests in the game without actually reading and understanding the text.

Quests are both the most entertaining and challenging way in which users of the Costello game are given points, and thus advance. A quest is a series of tasks in which the student must interact with the programmed characters in the game to help them overcome various problems. Without quests, players become bored with the game and feel that there is nothing to do except continually explore the world. Quests add a purpose to their exploration. The number of quests in Costello is continually expanding, but it still represents an area where more work is needed. Dedicated players (some of whom play the game more than ten hours every week) can finish all of Costello's quests within one term. The quests are all designed so that players must read and understand English in order to receive rewards in the game and some of them are very challenging puzzles, even for native speakers. At present there is a total of some 20 different quests within the game.

To put the Costello world descriptions and quests into the context of second language acquisition research, it

may be useful to recall Lightbown and Spada's (1993) summary of the arguments on comprehensible input:

1. Interactional modification makes input comprehensible.
2. Comprehensible input promotes language acquisition.

Therefore:

3. Interactional modification promotes acquisition. (p.30).

Although these arguments were largely designed for the spoken register, I believe that they can also apply equally to the role that written input plays in L2 acquisition. It is proposed that the extensive written texts provided by the Costello virtual reality provide input, that the interactional modification provided by tasks within the game makes the input comprehensible and that therefore the Costello virtual reality can play a real role in L2 acquisition.

The Builder

The builder is the application referred to in my previous paper that allows students to create their own areas within the Costello world without any knowledge of computer programming. A working version of this software has been available for some time, but it was decided that the interface needed simplification before students could usefully produce a substantial volume of written work within this environment. The software has thus been remodelled to work within a world-wide web browser, an interface that is already familiar to students. At the time of writing, students are working in groups, to produce their own original areas and quests that will later be added to the Costello world and enjoyed by their peers. It is already clear from the amount of work being produced that the chance to create something that will forever be part of the Costello virtual world is very motivating, even to the least gifted students (including a class of students who have already failed other English courses). However, the full results of a study of the student work produced within the Costello builder will be reserved for a future paper.

Part Two: Costello compared

In the second part of this paper, I will present the preliminary results of an ongoing experiment at Iwate Prefectural University involving the Costello virtual reality that examines two classes of students: an experimental group using Costello and a control group following a conventional extensive reading program. This study is now in its second year but at the time of writing, results are only available for the first complete year of the experiment. These results, as well as modifications made to the study in the second year, will be detailed and analysed. I will also present questionnaire data that comes from both of these classes.

Subjects

The subjects of the experiment were two classes of university students in their first year of study at Iwate Prefectural University. These students are non-English majors with a randomly assigned balance of students from four different faculties (Software and Information Science, Nursing, Social Welfare and Policy Studies). The course was a compulsory reading/writing course that was scheduled for a total of thirty 90 minute classes, once per week, split into two terms. This course was part of the general education requirements that all students at the university need to pass in order to graduate. The experimental group consisted of 45 students in both terms while the control group contained 42 students in the first term and 41 students in the second term.

Hypothesis

The hypothesis behind the experiment was somewhat modest. It was proposed that both groups would show

an advancement in reading ability and that this advancement should be of a comparable level. However, the experimental group should exhibit a stronger propensity to be reading for pleasure, shown both by time spent on task and reported enjoyment of the reading tasks. It is not claimed that the Costello virtual reality is a replacement for conventional extensive reading programs. Rather, it is being proposed that the game can be an equally effective tool that might have motivational benefits when compared with such a course. It may also offer an alternative approach for students who do not respond well to the conventional reading program or who show little enthusiasm for such work.

Method

Both classes were taught by the same instructor (myself) using the same reading textbook (LeBauer, 1997) in class. Listening tasks were also given in addition to the reading/writing tasks within the textbook itself. Class time was used in the same way in both groups except that the experimental group required two introductory classes in order to explain the Costello virtual reality. Some ten minutes of the Costello group classes were thereafter often required for updates and news about a system that was continually under development and expansion during the experimental period. In both groups, extensive reading work was allotted as homework and students were expected to either play the Costello MUD game (experimental group) or read from a graded reader of their choice (control group). These groups will henceforth be referred to as Costello group and Book group for clarity. Both groups were expected to submit a report through a form on the world-wide web in which they either detailed where they had gone and what they did in the virtual world (Costello group), or a summary in their own words of what they had read and their impressions about the content (Book group). These reports were graded by the instructor each week on a scale of 1-10 based on length and content (the main criteria being whether they had clearly expressed in their own English words what they wanted to say, and what they had experienced or read). In addition, the forms contained a number of questions relating to their perceptions of the week's work. These included perceived difficulty of the work, enjoyment, dictionary use and time spent on the task.

Administrative requirements meant that a separate grade was given at the end of each term since students received credits for each term's study. These grades were determined as follows:

	Costello Group	**Book Group**
Homework Report Scores	20%	20%
Classwork scores	50%	50%
Costello Game Scores	30%	
Book Pages Read		30%

A passing grade was 60%. The Costello game scores were a combination of three scores: a language score that reflected the number and length of successful commands they entered into the game, an exploring score that reflected how far they had travelled in the virtual world and an overall game score that reflected how much of the game they had completed. The 'book pages read' score was based on the number of pages the student read in their graded reader, multiplied by the level of the reader. Thus students could either read many pages in a low-level reader or fewer pages in a higher level reader to achieve the same score. Students were allowed to freely choose their reader and could change it at any time.

The evaluation of language ability for the purposes of both classroom administration and SLA research is a

subject that has attracted a great deal of debate and writing within the language-teaching community. (See, for example, Alderson and Beretta (1992), McNamara (1996), Bachman and Palmer (1996)). While a review of this literature is beyond the scope of this paper, it is clear that, providing that criteria are met for reliability and validity, there is no ideal testing solution for any situation. For the purposes of this study I decided to follow the arguments made by Bernhardt (1991) in favour of immediate recall protocols. To briefly summarise these arguments, she quotes Johnston's statement that they are the "most straightforward assessment of the result of the text-reader interaction" (Johnston, 1983, p.54) and goes on to give three arguments in favour of such tests:

"First, recall can show where a lack of grammar is interfering with the communication between text and reader, while not focusing a reader's attention on linguistic elements in texts. Second, generating recall data does not influence a reader's understanding of a text...Third, the measure is linked to an instructional strategy, thereby enabling it to maintain content validity." (Bernhardt, 1991, p.200-201).

Both the Costello group and the Book group in the current study were conducting reading tasks for homework that focussed on a general understanding and ability to process what they had read. Thus, the immediate recall test, which requires a basic understanding of a short text as a whole and in propositional units, was considered more satisfactory than other tests that often focus on aspects of reading ability such as grammatical awareness. Four tests were given to the students. The first was given in the very first class as a pre-test. The same test (using the same text) was given at the end of the first term and the end of the second term. Finally, a different and slightly longer text was also given at the end of the second term in order to assess the extent to which students were simply remembering the text content through repetition. The texts were excerpts from J.R.R. Tolkien's 'The Hobbit' and are presented in the appendix to this paper. The choice of text may seem to offer a bias to the Costello group but an examination of the texts chosen should reveal that only one word (troll) could actually be directly linked to a fantasy setting, and this word was not included in the assessment criteria.

Administration of the tests was conducted as follows: students were given copies of the text and were instructed to read it for a period of 10 minutes, after which they were required to stop looking at the text and write down in their native language (Japanese) what they could remember of the text content. They were given 5 minutes for the writing of the recall protocol. Students were not permitted to write anything during the reading part of the test. For the purposes of assessment, the texts were broken into propositional units (for example /they did not sing/ or tell stories/ that day/) and a point was given to the student if the unit content appeared in the protocol.

Results and discussion

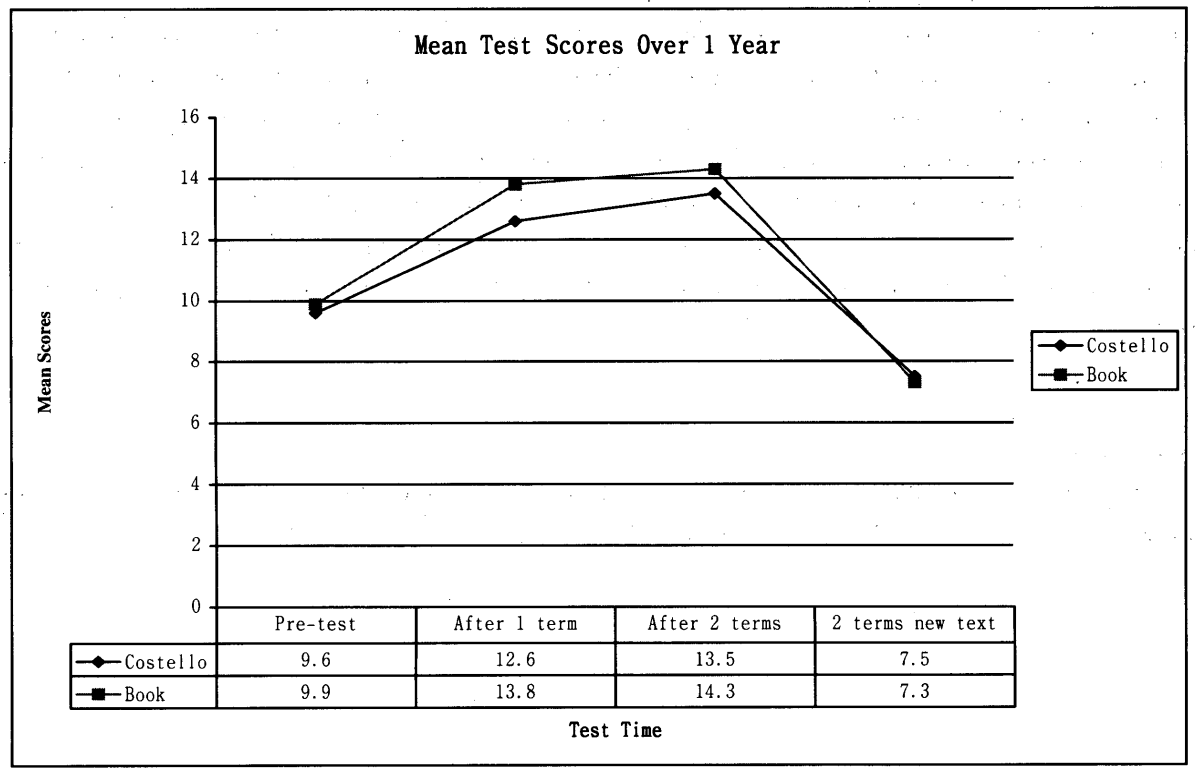
The results of the recall tests are shown below in both tabular and graphical format. In the tables 'Cos.' stands for Costello, 'Ss' for students, 'Pop.' for population, 'SD' for standard deviation and 'Std.' for standard.

After one term (4 months)

	Pre-test		Post-test 1	
	Cos (45 Ss)	**Book (42 Ss)**	**Cos (45 Ss)**	**Book (42 Ss)**
Mean	9.6	9.9	12.6	13.8
Pop.mean	8.2 to 11	8.3 to 11.5	10.8 to 14.4	12.2 to 15.4
SD	4.8	4.9	5.7	5.5
Std Error	0.7	0.8	0.9	0.8

After 2 terms (1 year)

	Pre-test 2 (Same text as 1st Term)		Post-test 3 (different & longer text)	
	Cos (45 Ss)	**Book (42 Ss)**	**Cos (45 Ss)**	**Book (42 Ss)**
Mean	13.5	14.3	7.5	7.3
Pop.mean	11.7 to 15.3	12.5 to 16.1	5.9 to 9.1	5.9 to 8.7
SD	5.7	5.6	5.3	4.6
Std Error	0.9	0.9	0.8	0.7



The data shows that the two classes were at almost exactly the same level before the course started but that by the end of one term the Book class had clearly made stronger gains than the Costello group, although both groups showed an improvement. This can perhaps be explained by the fact that the Costello group had a much longer learning curve than the Book group since they had to master the Costello game system, a process that is unfortunately somewhat complicated. Obviously there was a very small learning curve for the Book group, who simply had to read their books and write reports. This argument is supported by the fact that the two groups converge in the third repetition of the shorter text and we see that the Costello group actually performs marginally better than the Book group in the final test, in which a longer and slightly more difficult text (reflected by the lower scores), was used. Since the final test does not contain the possibility of students remembering content from the previous tests, it can be considered a fairer assessment of the two classes than the third test, in which students were seeing the same text for the third time. Thus the results would seem to uphold the hypothesis that the Costello virtual world provides a viable alternative to a conventional extensive reading program, since after one year the two groups achieved comparable results.

But what of the claim that students in the Costello group should be more motivated than the Book group? The following data is a summary of the students' responses to the weekly questionnaire on their homework reports. The data is taken from a sample of nine classes over the course of a year, since on a number of occasions full data was not submitted by students. 'How Difficult?' refers to the students' perceived difficulty of the task, 'How long?' is the reported amount of time they spent on task and 'Enjoyed' their perceived enjoyment of the task .

	How Difficult?		**How long?**	
	Cos (50 Ss)	**Book (43 Ss)**	**Cos (50 Ss)**	**Book (43 Ss)**
Mean Total hours	2.2	1.9	2.5 21.4	2.0 17.7
SD Total hours	0.7	0.6	1.3 12.2	0.9 8.2
Std Total hours	0.1	0.1	0.2 1.7	0.1 1.3
Pop. mean Total hours	2.0 to 2.4	1.7 to 2.1	2.1 to 2.9 18 to 24.8	1.8 to 2.2 15.1 to 20.3
Comment	Students marked from 1-5 1=Very Hard 5=Very Easy		Top score is reported hours/week. Bottom score is total reported hours after 9 classes.	

	Enjoyed?	
	Cos (50 Ss)	**Book (43 Ss)**
Mean Total hours	3.0	2.5
SD Total hours	0.7	0.8
Std Total hours	0.1	0.1
Pop. mean Total hours	2.8 to 3.2	2.3 to 2.7
Comment	Students marked from 1-5 1=Not at all 5=Very much	

This data does not in fact show a great difference between the two groups. The Costello group found the task slightly more difficult than the Book group but also spent longer playing the game and reported more enjoyment than the Book group. However the numbers are not statistically significant and it is impossible to conclude from this data that the game is indeed effective in providing a more motivating environment for English study.

Conclusion

The Costello virtual reality world is in a state of constant development. The results reported here show some potential for the system and some areas where it can be improved. For example, it is clear that the students are facing a difficulty in mastering and understanding the game in the initial stages. Thus the first stages of the game have been simplified greatly since the data taken here, and the documentation has been rewritten. Work has also started on a book that will provide an easier introduction to the game world. In terms of enjoyment, there are now many more quests than there were at the time of this study and there are also more quests that are suitable for beginning players. This is intended to reduce the learning curve and make the game enjoyable from the start. Of course the success of such efforts will only be determined by a continuing analysis of data and the experiments described here are being continued in the current academic year. Testing is also underway at Tohoku University, which is a higher level university and will provide insight into how different levels of student can affect the game's efficacy as a learning tool. It is the opinion of this writer that the potential for improvement that exists within Costello means that it will not be long before it shows itself to be a more effective tool than a conventional reading program. However, the truth (or falseness) of this claim will only become apparent with further research.

References

Alderson, J.C. and A. Beretta. (1992). *Evaluating Second Language Education.* Cambridge: Canbridge University Press.

Bachman, L.F. and A. Palmer. (1996). *Language Testing in Practice.* Oxford: Oxford University Press.

Bernhardt, E.B. (1991). *Reading Development in a Second Language.* New Jersey: Ablex.

Cohen, A.S. (2000),"Costello: Promoting Second Language Practice and Acquisition in a Text-based Virtual Reality Internet Game", Language and Culture, 2, p.93-125., Iwate, Japan: Iwate Prefectural University Center for Language and Culture Education and Research.

Ellis, R. (1994). *The Study of Second Language Acquisition.* Oxford: Oxford University Press.

Higgins, J. and T. Johns. (1984). *Computers in Language Learning.* London: Collins.

Johnston, P.H. (1983). *Reading Comprehension Assessment:* a Cognitive Basis. Newark, DE: International Reading Association. Cited in Bernhardt (1991).

Krashen, S.D. and T.D. Terrell. (1983). *The Natural Approach.* London: Prentice Hall.

LeBauer, R. (1997). *Journeys Reading 1.* Singapore: Prentice Hall Asia ELT.

Lightbown, P.M. and Spada, N. (1993), *How Languages are Learned,* Oxford: Oxford University Press.

McNamara, T. (1996). *Measuring Second Language Performance.* Harlow: Addison Wesley Longman.

Tolkien, J.R.R. (1937). *The Hobbit.* London: Harper Collins.

Appendix: Recall test texts.

1. Read the following text. After 10 minutes I will ask you to write down what you can remember.

They did not sing or tell stories that day, even though the weather improved; nor the next day, nor the day after. They had begun to feel that danger was not far away on either side. They camped under the stars, and their horses had more to eat than they had; for there was plenty of grass, but there was not much in their bags, even with what they had got from the trolls. One morning they forded a river at a wide shallow place full of the noise of stones and foam. The far bank was steep and slippery. When they got to the top of it, leading their ponies, they saw that the great mountains had marched down very near to them. Already they seemed only a day's easy journey from the feet of the nearest.

2. Read the following text. After 10 minutes I will ask you to write down what you can remember.

Tea-time had long gone by, and it seemed supper-time would soon do the same. There were moths fluttering about, and the light became very dim, for the moon had not risen. Bilbo's pony began to stumble over roots and stones. They came to the edge of a steep fall in the ground so suddenly that Gandalf's horse nearly slipped down the Slope.

"Here it is at last!" he called, and the others gathered round him and looked over the edge. They saw a valley far below. They could hear the voice of hurrying water in a rocky bed at the bottom the scent of trees was in the air and there was a light on the valley-side across the water.

Bilbo never forgot the way they slithered and slipped in the dusk down the steep zig-zag path into the secret valley of Rivendell. The air grew warmer as they got lower, and the smell of the pine-trees made him drowsy, so that every now and again he nodded and nearly fell off, or bumped his nose on the pony's neck.

Costello: Ongoing Experiments in an Internet Virtual Reality

Adrian Cohen

In an earlier paper, (Cohen, 2000) I introduced the concept of a text-based internet virtual reality (commonly called a MUD), as a pedagogical tool for use in an EFL setting. It was suggested that the interactive nature of such environments could provide motivation and opportunities for students to communicate in English, while the text basis of the virtual world would provide for comprehensible input according to the Krashen and Terrell (1983) model. The current paper has two purposes: to report on the current developmental status of the Costello virtual reality, and to detail the results of an initial experimental study comparing a class using the Costello RPG with a class using a conventional extensive reading program.